A Neural-based System for the Automatic Classification and Follow-up of Diabetic Retinopathies

A. Starita, A. Sperduti
Dipartimento di Informatica, Università di Pisa,
Corso Italia 40, I-56125 Pisa, Italy.
E-mail: perso@di.unipi.it

Abstract

We present a system able to provide and to simplify the automatic analysis of the diabetic retinopathies. The system stores and classifies the retinal images of diabetic subjects, allowing image filtering and anomalies localization by pattern's recognition methodologies combined with Neural Networks. The performance of the system has been quite improved by using committee techniques for the detection of retinal anomalies and with new algorithms specially designed to select the best classifiers to be inserted in the committee. The system has a simple, friendly but powerful interface supporting the medical personnel in the diagnostic process. To automate this process is particularly important studying the diabetic retinopathy for the large number of patients and the frequency of analysis necessary for following the evolution of the pathology.

1 Introduction

Diabetic patients show anomalies of the retina related to the state of gravity of the pathology, up to generate blindness. Then, the early diagnosis of these anomalies becomes important. In a Diabetic out-patients Department, it is routine to check the patients with a retinographic examination to get an early diagnosis of the damages to the retina and to evaluate the evolution of the illness with respect to the therapy. The standard retinographic examination currently consists of the manual analysis of 4 photos of the retina (2 fields of both eyes) to search for possible anomalies (microaneurysms, cotton wools, hard exudates, etc.; see Fig. 1)[6]. This method presents different disadvantages:

1. it is impossible to improve the quality of images, especially when these are very dark or with poor contrast, making difficult to recognise the

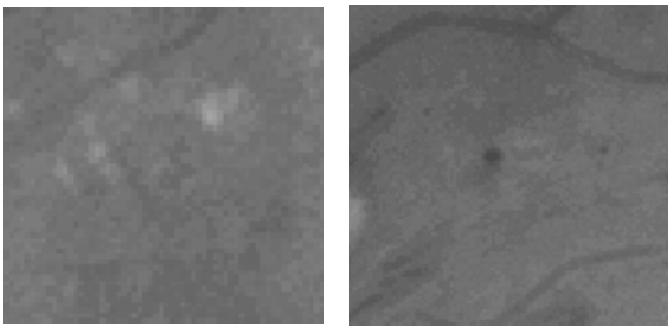


Figure 1: Examples of hard exudates (*left*) and microaneurysms (*right*).

characteristics of the anomalies;

2. to find the anomalies of the retinas is often a long and approximate process;

3. the high number of photos taken daily (up to 50), makes very difficult to reconstruct the clinical history of the patient;

4. between the inspection of the patient and the analysis of the patient's retina, it may pass a large number of days and this can be a problem when a rapid degenerative process is going on.

The DRA system (Diabetic Retina Analyser) [2], a system integrating patients management tools with intelligent image processing techniques, solves the previous problems, improving the clinical practice, simplifying the diagnosis, and also allowing the physicians to perform the follow-up of the patients. Moreover, the DRA system has a friendly but powerful interface that allows the visualization and/or handling of the images.

In Section 2 we briefly outline the general features of the DRA system. In Section 3 we describe the DRA's hybrid module which, combining image processing techniques with neural networks, performs the detection and localization of anomalies in the retina. Details on the use of committees of neural networks for the detection of anomalies are discussed in Section 4, while a comparison with a system based on neural networks facing a similar problem is presented in Section 5. Some information on the use of the system within the clinical environment are reported in Section 6. Conclusions are drawn in Section 7.

2 The DRA System

The DRA system (see Fig. 2) is structured in different modules (hardware and software) that interact each other as drawn in Fig. 3. The different modules of the system are described in the following:

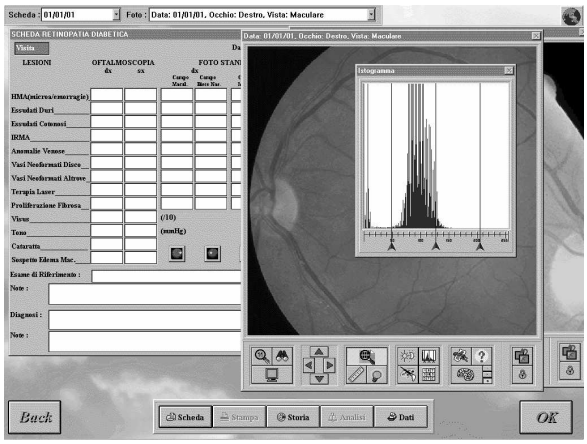


Figure 2: A particular of the system's graphical interface.

Retinographic Device: the device is an optical instrument which allows the physician to take photos of the retina (two for eye) by means of a special camera; afterward the photos are developed and stored on PhotoCDs.

Photo Acquisition: this module is used in order to store the photos; these are loaded from PhotoCDs and visualized on the monitor so that medical people can accept or discard them. Once associated to the patient the selected photos are stored in the Image Data Base.

Image Data Base: each image occupies a large portion of memory (about 1Mbytes/true photo, and about 300Kbytes/ compressed one). Then, it has been considered necessary also the use of a support of high capacity and renewable (because of the continuous growth up to 50 photos per day), like a CD-ROM writer or a Magnetic-Optic disk.

Patients Data Base: the Diabetes Unit has a data base containing all the anamnestic data relative to the patients. The system is able to connect to this data-base and to withdraw the whole necessary data.

History of the Photos: the physicians have a tool to visualize, in chronological sequence, all the images referred to a patient.

Photo Filer and Builder: these modules provide the physician with a set of tools to analyze the images, like the enlargement of peculiar zones or the measure of the distance between important biological features, and to implement various kinds of filtering (contrast, edges detection, stretching, etc.)[1]. It also allows the remodeling of the image in presence of degraded images (opaque, dark, blurred, etc.). Two kinds of automatic filters are implemented in this module: one transforms the

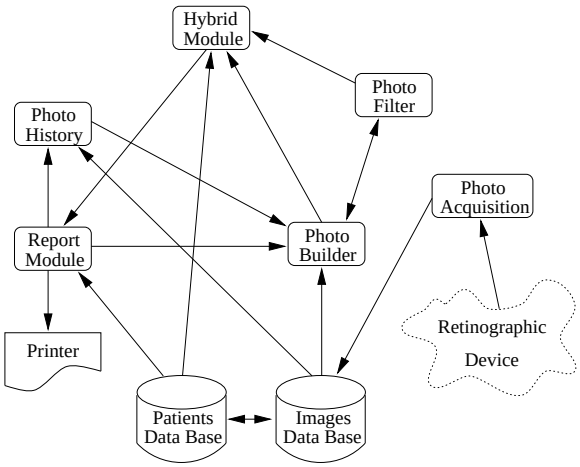


Figure 3: Structure of the DRA system.

images in a standard format (in respect of the level of brightness and contrast); the other, implements the analysis of the image highlighting all the anomalies of a selected type.

Hybrid Module: the Hybrid System allows the automatic detection and localization of anomalies of small and medium size.

Report: this module allows the physician to search the data, to compile the report on patients, writing down all the anomalies discovered during the inspection, and making easy the subsequent printing.

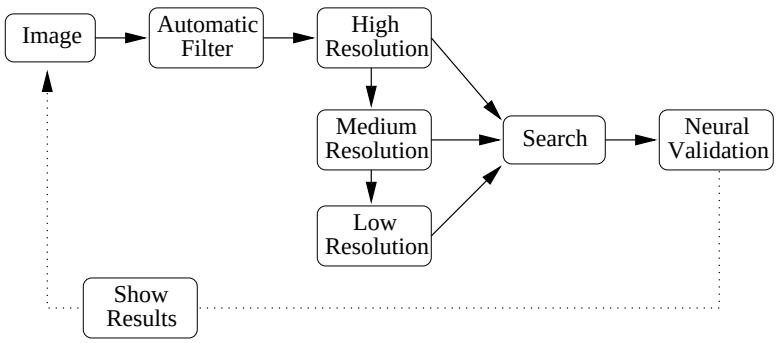
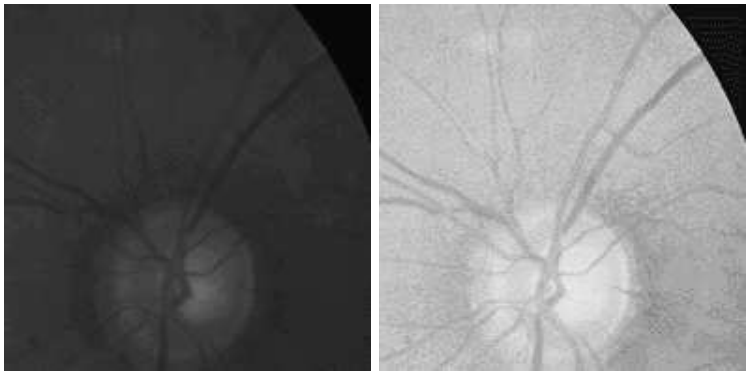


Figure 4: Functional scheme of the hybrid module.

3 Hybrid Module

The main feature of the hybrid module is to locate the anomalies which are clinically important, and to recognize them on the basis of their different shapes and colors (microaneurysm present themselves as little dark red spots, hard exudates are white spots with neat contours, etc.). The current version of the system can detect and localize microaneurysms and hard exudates[1]. This is obtained by selecting, through image processing techniques, zones of the retina which, with high probability, contain anomalies. The selected zones are then classified as true or false anomalies by specialized neural networks. In Fig. 4 the functional scheme of the hybrid module is reported. Specifically, the 512×768 pixels color image of the retina is pre-processed through a specially developed filter which on the basis of the brightness histogram, the green[2] component histogram, and on some statistics computed on a set of about 1,000 different images, implements an automatic stretch filter. The result is a normalized and enhanced version of the image which remarkably improves the quality of the image (see Fig. 5). The resolution of the image



a **b**

Figure 5: Detail of a retina before (**a**) and after (**b**) the application of the automatic filter.

is then changed according to the size of the anomalies which the physician is looking for. Images at higher resolutions allows the system to discover small size anomalies, and so on. In this way, the same *Search* and *Neural Validation* functions can be used when looking for anomalies of different sizes. The *Search* function works on the green component of the image and it exploits the observation that microaneurysms and hard exudates are local minima and maxima of the brightness, respectively. Specifically, the implementation of the *Search* function is based on a local thresholding procedure which using

[1] The last version of the system can also recognize cotton wools, however, this capability is currently under evaluation.

[2] We have selected the green component of the image since it is actually conveying all the relevant information about the anomalies.

a set of masks (up to a size of 5×5 pixels) selects the points in the image (actually, a 5×5 window around the point of interest) which are candidates to be microaneurysms and hard exudates. Usually, the *Search* function returns about 50-200 candidate points for each type of anomaly within 1-3 seconds on a 100MHz Pentium based Personal Computer. However, while for hard exudates only about 4% of the candidate points are actually false responses, this percentage greatly increases for microaneurysms. This is mainly due to the presence of veins which share several common signal features with microaneurysms. The best way to discriminate between this two classes of patterns is to use an adaptive technique. This is because the problem can only be defined through a set of positive and negative examples. On the basis of this consideration, the natural candidates to perform the task are feed-forward backpropagation networks. The task of the *Neural Validation* function is to remove from the list of candidate points the ones which are not anomalies. This is obtained by specialized neural networks, one committee of networks for each type of anomaly, trained on positive and negative examples of images returned by the *Search* function on a set of about 100 different images. Before to present the candidates (i.e., the 5×5 images returned by the *Search* function) to the networks for training (and testing), they are further processed by subtracting from all the pixels the value of the central pixel. In this way, the variation in the range of pixel values is normalized within a range of [-40,40] (see Fig. 6). This range is then mapped into the

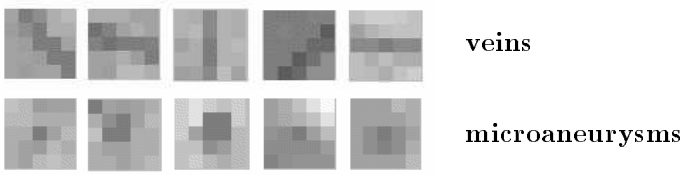


Figure 6: Examples of negative (veins) and positive (microaneurysms) patterns used in the training set of the neural network for microaneurysms.

[-1,1] interval for presentation to the neural networks. A mentioned above, for each type of anomaly, there is a committee of networks which provides to classify the input pattern. More details on the committees are given in the next section.

The anomalies detected by the hybrid module are then marked on the original image through a set of circles centered on the anomalies. An example of processing of the Hybrid Module is shown in Fig. 7.

4 Committee Algorithms

The problem faced by the *Neural Validation* module is not easy, and it is difficult to obtain good results by a single neural network. A way to improve the performance is to use committee of classifiers. For this reason, we have

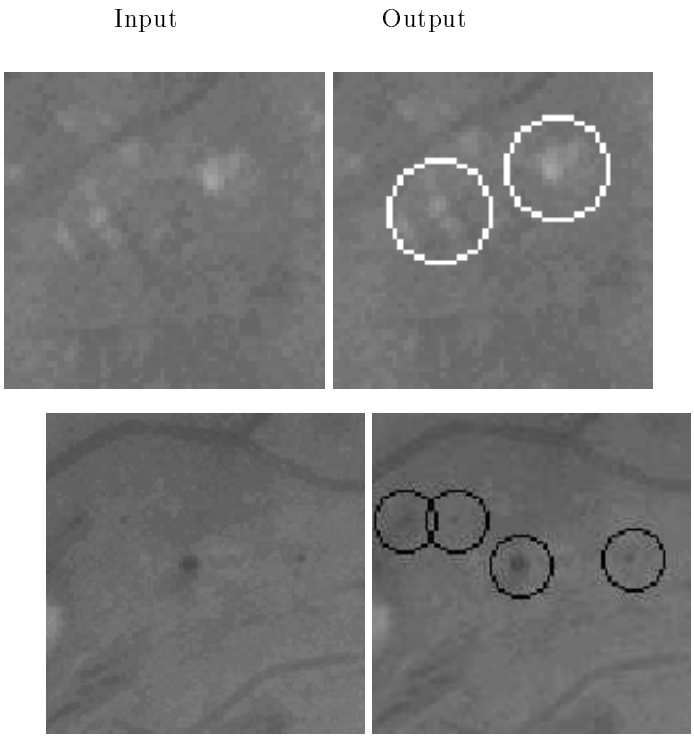


Figure 7: An example of results obtained by the Hybrid Module: results obtained on a portion of retina containing hard exudates *(top)*, and microaneurysms *(bottom)*.

experimented several committee algorithms [4]. Specifically, we have used the following combination techniques: Mean, Bayesian, OPM [7], Voting.

To improve the performance of the system, we have also devised new algorithms for the selection and combination of classifiers. Specifically, we show how to select the classifiers to be inserted in the final system using a greedy approach (*Greedy Selection* and *Pseudo-exhaustive Selection*), and how the sequential application of classifiers can improve the overall performance of the system (*Sequential Cooperation*).

4.1 New Selection Algorithms

It is well known that the error of a committee of classifiers depends both on the mean of the error of the classifiers and on the ambiguity between them [5]. Thus, in general, the insertion of a new classifier in the committee does not necessarily lead to an improvement of the performance. On the contrary, it is possible that the insertion of a new classifier (neural network) in the committee decreases the overall performance. To cope with this problem, we have developed two algorithms for the selection, from a pool of candidates, of the "best" classifiers to be inserted into the committee. For both algorithms, the Validation set is used for early stopping during training, while an additional independent set, called *Combination set*, is used to decide which candidate should be inserted into the committee.

4.1.1 Greedy Selection

The first algorithm is based on a greedy approach: the classifiers are sorted by increasing value of Mean Square Error (MSE) on the validation set; then the MSE of the committee obtained by inserting all classifiers starting from the first one to the i-th is computed on the combination set; finally, only the classifiers which caused a decrease in the MSE of the committees are inserted into the final committee:

Greedy Selection Algorithm
input: Classifiers $C_1,\ldots,C_N$, Validation Set, Combination Set;
output: Committee;

begin
for i=1 **to** N **do** Error_Val[i] $\leftarrow$ MSE(C_i,Validation Set);
Sort C_i according to increasing values of Error_Val;
for i=1 **to** N **do begin**
Committee $\leftarrow$ Committee $\cup$ C_i;
Error_Com[i] $\leftarrow$ MSE(Committee,Combination Set);
end
Committee $\leftarrow$ C_1;
for i=2 **to** N **do**
if(Error_Com[i]<Error_Com[i-1]) **then** Committee $\leftarrow$ Committee $\cup$ C_i;
end

4.1.2 Pseudo-exhaustive Selection

In this case the classifier with smallest MSE on the validation set is inserted into the committee and the MSE on the combination set is used as reference error for the committee. Then, in turn, all committees obtained by adding one of the classifiers are evaluated and the committee with smallest MSE on the combination set is selected. The algorithm stops when no committee generated in this way is better than the current one. Note that the generation of new committees is done with replacement, i.e., a classifier C_i which already is in the current committee can be selected again, leading to a duplication (which can also be multiple) in the new committee. This duplication must be avoided when the combination strategy is Bayesian or OPM, since the combination weight for each single classifier is automatically generated.

Of course, this algorithm is computationally much more expensive than the Greedy algorithm. Thus it is not practical to use it when the number of classifiers is huge. Finally, note that the algorithm is not really exhaustive, since no "backtracking" is allowed when a stop point is reached.

4.2 Sequential Cooperation

Usually, when combining several classifiers, the combination is performed in parallel. However, under specific conditions, an improvement on the performance can be obtained by sequentially combining the output of the classifiers.

Let consider a classification task with two classes ω_1 and ω_2. Moreover, let suppose we have available three classifiers, $C_i(\cdot)$, $i = 1, 2, 3$. These three classifiers can be combined in the following way

$$C_{seq}(\boldsymbol{x}) = \begin{cases} C_2(\boldsymbol{x}) & \text{if } C_1(\boldsymbol{x}) = \omega_1 \\ C_3(\boldsymbol{x}) & \text{if } C_1(\boldsymbol{x}) = \omega_2 \end{cases} \tag{1}$$

The resulting classifier $C_{seq}(\cdot)$ will show an improved performance over the single classifiers $C_i(\cdot)$ under specific conditions on the classifiers's sensitivity. Remembering that the sensitivity of a classifier C_i with respect to class ω_j is defined as $S_i(\omega_j) = P(C_i(\boldsymbol{x}) = \omega_j | \boldsymbol{x} \in \omega_j)$, it is possible to derive necessary conditions for C_{seq} to be superior in performance to each single C_i.

In fact, under the hypothesis that classifier C_1 is independent with respect to C_2 and C_3, and applying Bayes rule, the sensitivity of C_{seq} with respect to ω_1 can be expressed as

$$S_{seq}(\omega_1) = S_1(\omega_1)S_2(\omega_1) + (1 - S_1(\omega_1))S_3(\omega_1), \tag{2}$$

while the sensitivity with respect to ω_2 is

$$S_{seq}(\omega_2) = S_1(\omega_2)S_3(\omega_2) + (1 - S_1(\omega_2))S_2(\omega_2). \tag{3}$$

Finally, the performance of C_{seq} can be expressed using $S_{seq}(\omega_1)$ and $S_{seq}(\omega_2)$

$$Performance(C_{seq}) = S_{seq}(\omega_1)P(\boldsymbol{x} \in \omega_1) + S_{seq}(\omega_2)P(\boldsymbol{x} \in \omega_2). \tag{4}$$

In order for C_{seq} to be superior with respect to each single classifiers C_i, we must have

$$Performance(C_{seq}) > Performance(C_i)\ i = 1, 2, 3. \tag{5}$$

By using equations (4) and (2-3), it turns out that the following conditions have to be satisfied:

Superiority with respect to C_1 (6)

$$\begin{aligned} S_1(\omega_1)[S_2(\omega_1) - S_3(\omega_1) - 1] \quad > \quad & \frac{P(\omega_1) - 1}{P(\omega_1)}(S_1(\omega_2)[S_3(\omega_2) - S_2(\omega_2) - 1] \\ & + S_2(\omega_2)) - S_3(\omega_1) \end{aligned}$$

Superiority with respect to C_2 (7)

$$S_1(\omega_2)[S_3(\omega_2) - S_2(\omega_2)] \quad > \quad \frac{P(\omega_1)}{1 - P(\omega_1)}(S_1(\omega_1) - 1)[S_3(\omega_1) - S_2(\omega_1)]$$

Superiority with respect to C_3 (8)

$$S_1(\omega_1)[S_2(\omega_1) - S_3(\omega_1)] \quad > \quad \frac{1 - P(\omega_1)}{P(\omega_1)}(S_1(\omega_2) - 1)[S_2(\omega_2) - S_3(\omega_2)]$$

It is not difficult to see that, under the conditions of reasonably balanced a priori probabilities for the two classes ω_1 and ω_2 and given a reasonably good classifier C_1, conditions (7) and (8) are satisfied by the joint events *i)* $S_2(\omega_1) > S_3(\omega_1)$ and *ii)* $S_3(\omega_2) > S_2(\omega_2)$. On the contrary, it is more difficult to understand condition (6), which relates the sensitivity of the three classifiers in a more complex fashion.

4.3 Experimental Results

In this section we report the experimental results obtained by applying standard committee techniques (Mean, Bayesian, OPM, Voting) without selection of classifiers, with selection by our algorithms (Greedy and Pseudo-exhaustive Selection), and Sequential Cooperation.

The images for microaneurysms (hard exudates), i.e. 5×5 images generated by the *Search* module, are split in four sets: a Training set of 1200 (500) images, a Validation set (for early stopping) of 576 (151) images, a Combination set of 276 images[3], and a Test set of 600 (350) images.

The number of neural networks (classifiers) used in the experiment was 10 for hard exudates and 30 for microaneurysms. They differed both in number of hidden units and in the learning parameters used during learning. For hard exudates the best network was able to obtain an error rate of 3.2. By combining the 10 networks together we obtained the following results for the error rate: Mean, 2.9; Bayesian, 2.6; OPM, 2.0; Voting, 2.0. Because of the good results, no experiment with the selection algorithms and the Sequential Cooperation algorithm was performed. Results for microaneurysms are shown

[3]The selection algorithms have been experimented only over microaneurysms.

Table 1: Microaneurysms: experimental results for different combination algorithms, with (I = Invariants) or without invariants. NN is the best neural network over 30 different networks.

	NN	NN_I	Mean	$Mean_I$	Bayes	$Bayes_I$
% Error	12.7	11.4	11.0	10.2	11.0	9.4
	OPM	OPM_I	Voting	$Voting_I$		
% Error	10.6	9.9	10.3	9.0		

Table 2: Microaneurysms: experimental results, using invariants, for different combination and selection (G = Greedy, E = Exhaustive) algorithms. Results for the Sequential Cooperation algorithm are reported as well (Seq).

	$Mean_G$	$Mean_E$	$Bayes_G$	$Bayes_E$	OPM_G	OPM_E
MSE	0.0335	0.0334	0.0342	0.0341	0.0342	0.0327
% Error	8.9	9.0	9.0	9.0	9.1	8.3
	$Voting_G$	$Voting_E$	Seq			
MSE	-	-	0.0371			
% Error	9.1	8.9	7.8			

in Table (1). The table reports also results obtained by using invariants. The basic idea is to transform the input using a transformation for which we want the classification to be invariant (we used clockwise rotations of 90, 180, and 270 degrees, horizontal and vertical reflexive projection), and then to evaluate the output of the networks for these transformed images. The obtained outputs are then used within the committee, as if they were outputs of additional new classifiers[4]. It can be noted that invariants increase the performance of the committee. Results obtained for the selection algorithms and the Sequential Cooperation algorithms are reported in Table (2). It is evident that the performance increases independently of the combination strategy, thus validating the quality of the proposed selection algorithms. The best result is obtained by the Sequential Cooperation algorithm.

We can conclude that the use of committees in the automatic detection of microaneurysms and hard exudates leads to a significant improvement of the performance of the system. Moreover, we have shown that it is possible to further improve the performance by using invariants, new selection algorithms, and sequential cooperation among single classifiers.

[4] Note that the response time of a committee using invariants is slowed down of a factor 5, i.e., the number of transformations, with respect to the same committee without invariants.

5 Related Work

The use of neural networks for the classification of diabetic retinopathy is explored in [3] where a neural network with an input window of 20×20 pixels is moved across the image. The main difference with respect to our work is that the task in [3] is to automatically classify the images, while the aim of our system is the localization of small anomalies to support early diagnosis of retina pathologies. It must be noted that our task is more difficult, since images which present both small and large size anomalies can easily be classified on the basis of the large size anomalies only. Nevertheless, our hybrid system is considerably faster since the neural networks both are smaller (20 versus 200 hidden nodes) and need to be applied only on a small fraction of the pixels of the images. Concerning the classification performance, it is not possible to make a direct comparison because of the different formulation of the classification task and to the different sets of images used for the training of the networks, however the performances of our networks are comparable to the results obtained by the networks developed in [3]. The physicians using our system[5] are very satisfied about the speed and performance of the system.

6 Validation of the DRA System

An experimental clinical screening of patients affected by diabetic retinopathy was performed by using the DRA System [8]. The screening was carried on at the Metabolic Diseases Department of University of Pisa, and lasted six months, during which the physicians have examined 2533 subjects. The sample of 2533 subjects was made of 1314 males and 1219 females, with mean age 55.1 ($\pm$15.5), suffering diabetes from 12.9 ($\pm$10.2) years. A so long period of illness has been chosen because it is known a positive correlation between the retinopathy and the length of illness. The obtained results have been in accordance with the data on retinopathy, known at European level. The use of the DRA System during the screening has allowed to improve the quality of images, to single out early the microaneurysms and the hard exudates present on the retina, so to prevent immediately the possible blindness of the patients. Moreover, the facility to recall and to easily store the collected images, relative to a patient, in the clinical database has allowed the follow up of the subjects and their therapy.

7 Conclusion

The DRA system is currently in use at the Diabetes Unit of the University of Pisa, which as an archive of about 10,000 photos of retinae and a data base of

[5] The system is actually functioning at the Reference Regional Centre for Diabetes Mellitus of Tuscany in Italy.

about 3,000 patients, running on an Unix server on which Oracle is the database manager. The DRA system runs on a Pentium 100MHz with 32Mbyte of RAM. The system as an open architecture so that in the future it can be expanded, adding new modules such as modules for statistical analysis of anamnestic data and search for new relationships between clinical and statistical parameters. The next goal will be the development of an expert tutorial system for diabetic retinopathy.

As a final comment we would like to stress that the Hybrid Module implemented within the DRA system is able to detect and to localize anomalies very fast and in a reliable way. Moreover, since standard image processing techniques are used, it is very robust to noise. These results were obtained by combining the high computational speed of the *Search* function with the high discriminative capabilities of the *Neural Validation* function. A similar result would have been impossible to reach by only using neural networks over the whole image because of the high computational cost of such approach.

References

[1] G.A. Baxes. *Digital Image Processing.* Wiley, New York, 1994.

[2] E. Casi, A. Ceravola, R. Cionini, A. Sperduti, and A. Starita. Diabetic Retina Analyzer, *Proceedings of the 18th Annual International Conference of IEEE Eng. in Med. and Biol. Soc.* Vol 4, pp. 282-283, Amsterdam, 1996.

[3] G. G. Gardner and D. Keating and T. H. Williamson and A. T. Elliott. Detection of Diabetic Retinopathy Using Neural Network Analysis of fundus Images. In *Proc. 2nd International Conference on Neural Networks and Expert Systems in Medicine and Healthcare*, Plymouth, UK, 1996.

[4] S. Haykin. Neural Networks: A Comprehensive Foundation. Second Edition, Prentice Hall, 1998.

[5] A. Krogh and P. Sollich. Statistical Mechanics of Ensemble Learning. *Physical Review E*, Vol. 55, pp. 811-825, 1997.

[6] R.J. Olk and Carol M.Lee. *La Retinopatia Diabetica*, Vol. I, II. Mediserve, 1993.

[7] K. Tumer and J. Ghosh. Error Correlation and Error Reduction in Ensemble Classifiers. *Connection Science*, Vol. 8, n. 3&4, pp. 385-404, 1996.

[8] S. Viti. Metodologia Computerizzata per l'Elaborazione delle Immagini del Fundus Oculi per lo Studio della Retinopatia Diabetica. "Tesi di Specializzazione", University of Pisa, 1997. In italian.